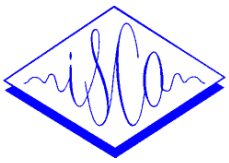

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

THE RHYTHM RULES IN JAPANESE BASED ON THE CENTERS OF ENERGY GRAVITY OF VOWELS

Masayo Katoh, Shin'ichiro Hashimoto

SECOM Intelligent Systems Laboratory, SECOM Co., Ltd.
6-1-1, Sakaemachi, Tachikawa, Tokyo, 190 JAPAN

Abstract

This paper proposes some new rules of rhythm in Japanese, based on *the Center of Energy Gravity of Vowels*: CEGV. These rules have risen from isochrony of mora, which linguists consider a characteristic feature of Japanese. Assuming the CEGV as the timing point of rhythm, and the duration between CEGV's(D_{cegv}) as the parameter to determine the rhythm, we can form the Japanese rhythm simply by using only 29 rules (*the first approximate rule-set)*. The isochrony of more than two moras is said to be another Japanese feature. Based on this, an additional rhythm rule (*the second approximate rule)* is proposed to prevent rhythm disturbances due to some exceptional combinations of the rules in the first approximate rule-set.

10.21437/ICSLP.1992-135

1 Introduction

It must be possible to describe the Japanese speech rhythm through simple rules, since even young Japanese children can speak the language very fluently. Also, native speakers can read aloud any new text, with natural rhythm.

Recently, practical engineering approaches have been to assign many complicated duration rules. Takeda et al.(1989) formulated a statistical duration control model through factor analysis of a set of 5240 words. Since the factors which define the duration are inter-related in a complicated way, Nakajima et al.(1989) avoided the use of duration rules and modified clustering matrix quantization techniques to generate a synthesis unit. Nevertheless, the synthesized voice is still unnatural. It may be that the fundamentals of the duration rules are yet to be discovered.

In this paper, we present a new Japanese rhythm theory which arises from reconsidering the utterance phenomena. Based on this theory, two sets of duration rules are proposed.

2 A new Japanese rhythm theory

We propose to characterize Japanese rhythm by the duration between the centers of energy gravity of vowels(D_{cegv}), due to the following observations.

(1)Speaking Japanese very slowly, it is easy to recognize that the timing point of the rhythm is located in the vowels and not in the consonants.

(2)Considering the physical structure of the ear, people may not listen to the waveform itself, but its energy profile.

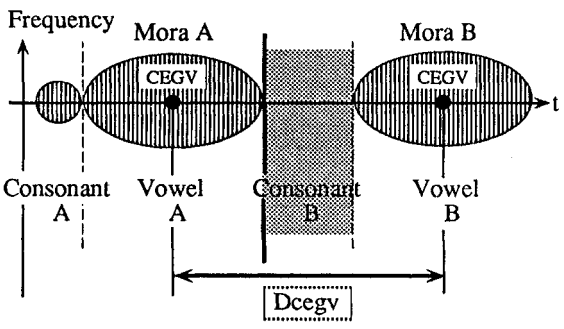

Figure 1: The center of energy gravity of vowels(CEGV) and the duration between CEGV's(D_{cegv})

Therefore, we propose to designate the center of energy gravity of vowels(CEGV) as the timing point of the rhythm(Fig.1).

Since linguists generally consider the Japanese moras to be isochronal, we assume the D_{cegv} to be essentially isochronal in normal speech. Such an isochrony is disturbed mostly by the difficulty of utterance, not of the vowels, but of the consonant in between the vowels. The difficulty of utterance of the consonants is caused by the vocal organ's structural restrictions. The single mora isochrony is used here to develop the first approximate rule-set, which will be described later. Furthermore, linguists also propose the isochrony of a set of two moras. This aspect is used to enhance the synthesis by adding a second approximate rule to the first approximate rule-set.

3 Phonemic symbol sequence vs. D_{cegv}

In this section, we investigate the hypothesis that the D_{cegv} is decided not by the vowels but the consonant in between the CEGV's.

The phonemic environmental differences, which may effect the D_{cegv}, are analyzed for the following cases: (1) Different preceding vowels in VCV (Vowel, Consonant, Vowel) (2) Different succeeding vowels in VCV (3) Nasals and diphthongs (4) Different preceding consonants in CVCV (5) The consonants in VCV (6) Choked sounds in VCV (7) Devoicing sounds in VCV

The speech material, consisting of a seven mora no-sense word, *"ko-ba-ba-me-N-ka-i"*, uttered 10 times by a female native speaker. This word was embedded in a constant carrier sentence, *So-re-wa — de-su*("That is -"). The phoneme of the second or the third mora(-ba-ba-) is changed to generate the various cases listed above. The D_{cegv} analysis is done between the second and third CEGV. The average and the standard deviation of the D_{cegv} are calculated after normalization of the tempo(7moras/sec).

Next, the D_{cegv} at the beginning and the end of the seven mora no-sense word are investigated. The pitch differences, at the beginning of the word, are also analyzed. The D_{cegv} between the first and second, or the sixth and seventh moras are analyzed. The average and the standard deviation are calculated as before.

4 Analyses results

Results of above analyses are presented in Figs.2~9.

A time duration of less than 20ms is ignored as insignificant. The results demonstrate that *the D_{cegv} is mainly decided only by the consonant type in VCV*(Figs.2~6). The choked sounds add only a constant duration 100ms to original consonant(Fig.7).

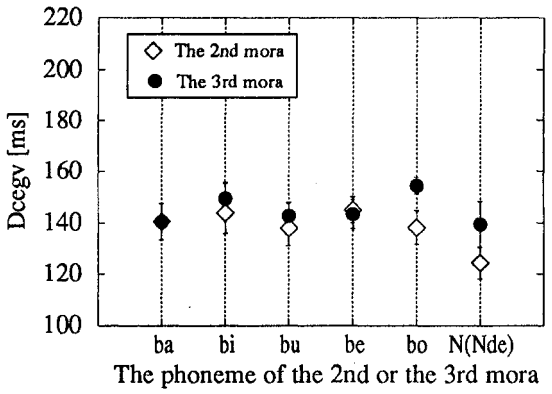

Figure 2: D_{cegv} vs. the difference of vowels

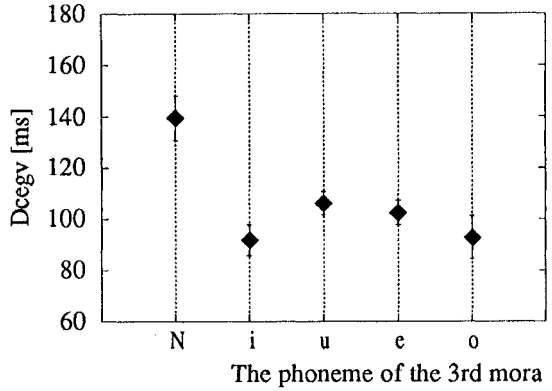

Figure 3: D_{cegv} vs. the difference of vowels(no-consonant)

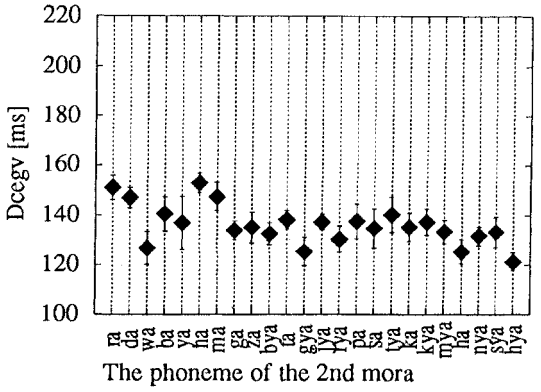

Figure 4: D_{cegv} vs. the difference of the 2nd consonants

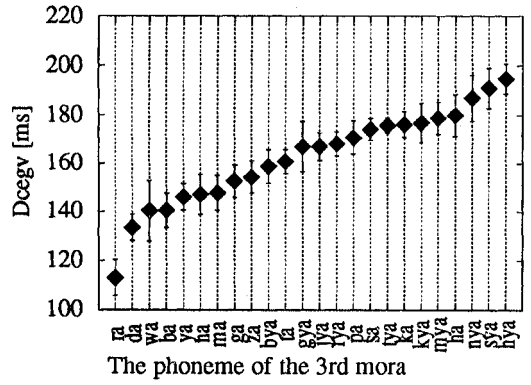

Figure 5: D_{cegv} vs. the difference of the 3rd consonants

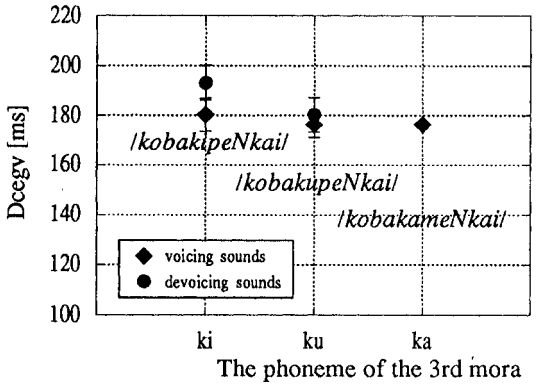

Figure 6: D_{cegv} of the devoicing vs. the voicing sounds

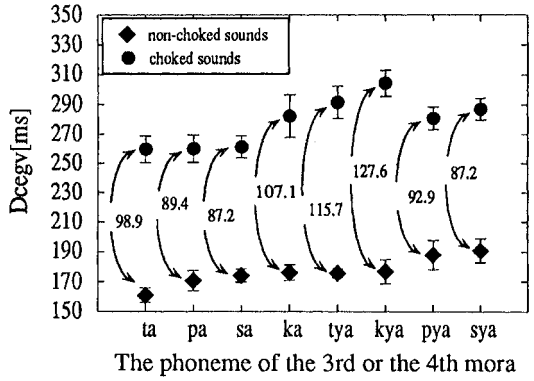

Figure 7: D_{cegv} of the choked vs. the non-choked sounds

At the beginning of the word, the D_{cegv} gets shorter as compared to the middle by about 20ms, except for the six explosive palatalized consonants: by, ty, zy, ky, py and gy (Fig.8). In addition, the D_{cegv} at the beginning is not much affected by the pitch differences. Compared with the beginning, the D_{cegv} at the end can be longer or shorter, and is not very critical(Fig.9).

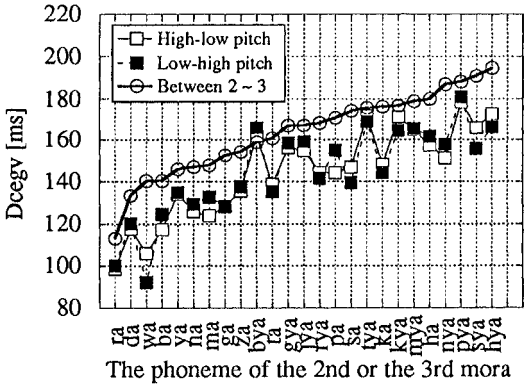

Figure 8: D_{cegv} at the beginning of the word(with pitch difference)

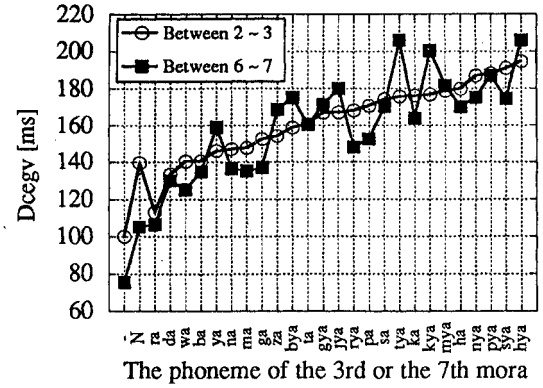

Figure 9: D_{cegv} at the end of the word

5 The first approximate rule-set

The first approximate rule-set is proposed based on these results. It consists of 29-rhythm-rules as in Table 1. Using this rule-set, it is possible to define all D_{cegv}s in Japanese.

6 the second approximate rule

Although the first approximate rule-set can characterize all Japanese VCV segments, the synthesized speech produced with this rule-set has an unnatural rhythm in some special cases. Fig.9 shows two such examples.

Assuming that the speech tempo is 7moras/sec, one standard D_{cegv} is about 145ms. If more than two shorter D_{cegv}s (for instance of 100ms each) are together, the duration around this part gets shorter by 90ms as compared with other parts. On the other hand, if longer D_{cegv}s (for example of 190ms each) are together, the duration gets longer by 90ms. In both these

Table 1: The first approximate rule-set

consonant	D_{CEGV}	consonant	D_{CEGV}
no consonant	98	jy	167
syllabic nasal	140	ry	168
r	113	p	171
d	134	s	174
w	140	ty	176
b	141	k	176
y	146	ky	177
n	147	my	179
m	148	h	180
g	153	ny	187
z	154	py	188
by	159	sy	191
t	161	hy	195
gy	167		

♣ choked: each consonant + 100[ms]
♣ at the begging(except 6 consonants):
each consonant -20[ms]

cases, the disturbance of the rhythm is clearly perceived. In order to diminate this defect, it is necessary to take into account the second rhythm feature of the isochrony of a set of two moras. Therefore, the following second approximate rule is proposed(Fig.10). If two consecutive (sometimes three) moras are too short or too long with respect to isochrony, then the two or three corresponding $D_{cegv}s$ are modified to satisfy isochrony.

(1)Let S be sum of n $D_{cegv}s(n = 2or3)$ using the first approximate rule-set.

(2)If S is in the interval $((145 \pm 20)n)$ no correction is needed.

Otherwise, the following correction is applied:

$$New\_D_{cegv} = \frac{Old\_D_{cegv}}{S} \times 145n$$

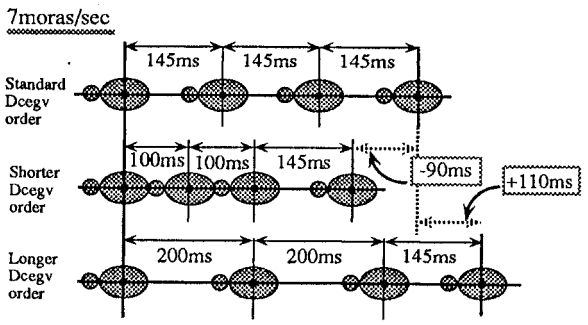

Figure 10: Examples of special cases

The speech synthesized with the first and second approximate rule-sets has a more natural rhythm than the one with the first approximate rule-set only.

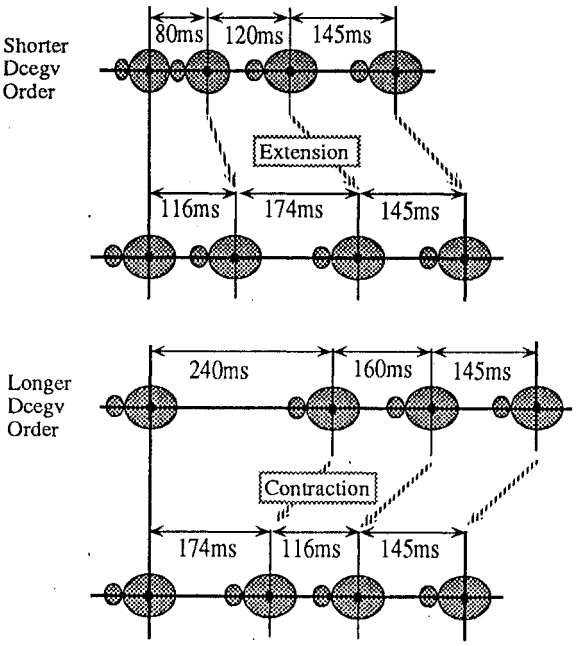

Figure 11: The second approximate rule

7 Conclusion

A new Japanese rhythm theory is proposed: the center of energy gravity of vowels(CEGV) is the timing point of rhythm and the duration between CEGV's(D_{cegv}) defines the Japanese rhythm. Based on this theory, the first and second approximate rule-sets are proposed. Though these rule-sets contain only 30 rules, the synthesized speech is quite natural.

References

[1]M.Katoh and S.Hashimoto, " The Rhythm Rules in Japanese Based on the Center of Energy Gravity of Vowels," IEICE Technical Report, SP92-11, pp.33-40, 1992.
[2]K.Takeda,Y.Sagisaka and H.kuwabara, "On sentence-level factors governing segmental duration in Japanese," J.Acoust.Soc.Am.86(6), pp.2081-2087, 1989-12.
[3]S.Nakajima and H.Hamada, "Speech Synthesis Method Based on Context Oriented Clustering,"Trans.IEICE PartD-2, Vol.J72-D-2, No.8, pp.1174-1179, 1989-8.